DOCUMENT RESUME

ED 240 125 TM 830 721

AUTHOR Urman, Harold; And Others
TITLE The Effect of Test-Wiseness Training on the Achievement of Third and Fifth Grade Students.
PUB DATE Apr 83
NOTE 31p.; Presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, April 12-14, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Gains; Elementary Education; Intentional Learning; *Mathematics Achievement; Pretests Posttests; Racial Differences; *Reading Achievement; *Scores; Sex Differences; Test Coaching; *Test Wiseness

ABSTRACT

The objective of the present research was to investigate the effects of test-wiseness training on achievement. The study involved 208 Black, White and Hispanic students in grades three and five. An experimental training session was employed to test whether four aspects of test-wiseness (following directions, guessing, answer changing, and use of time) would interact with race and sex in math and reading achievement using a post- and delayed post-test. For the fifth grade, the results indicated that students in the training groups scored higher than the controls in math on both the post- and delayed post-test. No training effects were significant at the third grade, nor were any of the interactions between training and race or training and sex at either the third or fifth grade. Recommendations for futher research are offered. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED240125

THE EFFECT OF TEST-WISENESS TRAINING ON THE ACHIEVEMENT OF THIRD AND FIFTH GRADE STUDENTS

HAROLD URMAN

JERI BENSON

DENNIS HOCEVAR

UNIVERSITY OF SOUTHERN CALIFORNIA

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. Urman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at the Annual Meeting of the National Council on Measurement in Education Montreal, April, 1983

Printed in the United States

TM 830 721

ABSTRACT

The objective of the present research was to investigate the effects of test-wiseness training on achievement. The study involved 208 Black, White and Hispanic students in grades three and five. An experimental training session was employed to test whether four aspects of test-wiseness (following directions, guessing, answer changing and use of time) would interact with race and sex in math and reading achievement using a post and delayed posttest. For the fifth grade, the results indicated that students in the training groups scored higher than the controls in math on both the post and delayed posttest. No training effects were significant at the third grade, nor were any of the interactions between training and race or training and sex at either the third or fifth grade. Recommendations for future research were offered.

The Effect of Test-wiseness Training on the Achievement of Third and Fifth Grade Students

There has been a growing dependency on using group administered tests, either standardized, competency-based or criterion-referenced to classify, screen or promote school children. Some children, especially those from minority groups (Dreisbach & Keogh, 1982) may be ill-prepared for taking tests, not because of any lack of content knowledge, but because of a 'handicap' in test-wiseness skills. For the purpose of this study, test-wiseness is defined as "a student's capacity to utilize the characteristics and formats of the test and/or test-taking situation to receive a high score" (Millman, Bishop & Ebel, 1965 p. 707). More importantly, test-wiseness is thought to be conceptually independent of the examinee's knowledge of the subject matter (Slakter, Koehler & Hampton, 1970a, b). Therefore, test-wiseness is considered to be an additional source of error variance in a test score (Ebel, 1965; Svanley, 1971; Thorndike, 1951).

A seminal work in the area of test-wiseness was the taxonomy developed by Millman, Bishop and Ebel (1965). Millman et al. conceptualized test-wiseness in terms of the characteristics of the test, (e.g., the correct answer may be longer, more familiar, more plausible, etc.) and the charcter-

istics of the examinees, (e.g., good test-takers may make better use of their time, employ appropriate guessing strategies, follow directions, eliminate incorrect answers for reasons other than their knowledge of the subject matter, etc.).

As indicated by a recent literature review (Sarnacki, 1979) considerable attention has been given to the construct of test-wiseness. To date, the research points to several conclusions. First, the construct of test-wiseness is complex, including dozens of different behavioral and mental components. Second, lack of test-wiseness can penalize certain students (Benson, 1980; Callenbach, 1973; Crehan, Koehler & Slakter, 1974; Crocker & Benson, 1980; McMorris, Brown, Snyder, Pyrczak, 1972; Mercer, 1979; Slakter, 1968a, b; Slakter et al., 1970a, b; Woodley, 1973, 1975). Third, the bias against students who are not test-wise extends to standardized tests which are professionally developed such as the Metropolitan Readiness Test, Comprehensive Test of Basic Skills, Metropolitan Achievement Test, Stanford Reading Test etc. (Benson, Urman & Hocevar, 1982; Metfessel & Sax, 1958)

Sarnacki (1979) suggested in his literature review that there are two alternatives available to minimize or eliminate the effects of test-wiseness. One possible approach would be to provide an instructional program for test makers on test-taking, test construction, and test-wiseness. Indi-

viduals would be trained in how to eliminate secondary cues in tests. Unfortunately, such a program would not be a guarantee that secondary cues would not appear in the future, nor is test-wiseness totally based on the detection of secondary cues.

The second and probably most viable solution would be to identify those individuals who are low in test-wiseness skills and train them in these specific skills. Although this task would not be easy, measures have already been developed to identify some of the skills that students lack (Diamond and Evans, 1972; Gibb, 1964; Slakter, et al., 1970b). Thus, if differences associated with a lack of test-wiseness skill could be reduced, then the error variance attributed to this variable could be minimized across subgroups of students.

To date, numerous methods have been used to teach or manipulate different aspects of test-wiseness (Bajtelsmit, 1977; Callenbach, 1973; Dreisbach & Keough, 1982; Flynn & Anderson, 1977; Ford & Weener, 1980; Gaines & Jongsma, 1974; Gibb, 1964; Gross, 1976; Kalechstein, Kalechstein & Doctor, 1981; Langer, Work & Johnson, 1973; McMorris et al., 1972; Moore, Schutz & Baker; 1966; Moore, 1971; Morse & Morse, 1980; Oakland, 1972; Omvig, 1971; Pike, 1978; 1980; Powers & Sabers, 1981; Schmitt & Crocker, 1981; Slakter et al., 1970 a, b; Wahstrom & Boarma, 1968; Woodley, 1973; 1975; Woo, Hocevar & Benson, 1983). The great diversity of the numer-

ous operationalizations of test-wiseness manipulations is readily apparent, ranging from nonobtrusive, short term manipulations to more general, multifaceted training curricula. Examples of the former are Schmitt & Crocker's (1981) manipulation of test instructions, and Woo et al. (1983) manipulation of the placement of the correct answer. Examples of the latter are described below. Also apparent in the literature is the diversity of subjects (preschool to adult) and outcome measures (test-wiseness skills, aptitude, standardized and teacher-made achievement measures) employed to evaluate the efficacy of test-wiseness training.

Of particular interest in the present study are interventions of a general nature which use elementary school children as subjects. Two such studies (Callenbach, 1973; Gaines & Jongma, 1974) are noteworthy. Using 48 students in two second grade classrooms, Callenbach instructed students over a four week period for a total of four hours in general testing skills using content that was nonverbal (geometric figures, picture and math computation problems). When controlling for verbal ability, significant differences were found in students scores on the Stanford Achievement Test administered one week after treatment. This difference was still significant on a delayed post-test given four months later using a parallel form of the Stanford Achievement Test. In a similar study, Gaines and Jongma (1974) constructed a unit on general test-taking tips for fifth grade

students. The group that received the unit of instruction scored significantly higher than the control group on four subtests of the California Test of Basic Skills.

While there are a substantial number of empirical investigations of test-wiseness training, there is a notable lack of research that deals with minority populations. This is a significant shortcoming in that educators have provided considerable documentation of the poor performance of minority groups on standardized achievement tests, and a purported explanation for this finding is that minority groups are not test-wise (Anastasi & Cordova, 1953; Eels, David, Havinghurst, Herrick & Tyler, 1951; Goslin, Epstein & Hallach, 1965; Kalechstein, et al., 1981; Mercer & Brown, 1973; Ortar, 1960). Of the 26 test-wiseness intervention studies cited above, only three studies used non-white subjects as the target population (Oakland, 1972; Kalechstein et al., 1981; Dreisbach & Keogh, 1982).

The first published analysis of test-wiseness training in a non-white population was a study conducted by Oakland (1972). While Oakland did not statistically analyze ethnic differences, his sample included both Black and Hispanic preschoolers in a Head Start program. His intervention consisted of a multi-faceted curriculum specifically aimed at abilities which were hypothesized to be prerequisite for taking readiness tests. After six hours of instruction over six weeks, Oakland found that an experimental group outper-

formed a control group on the total score of the Metropolitan Readiness Test administered immediately after completion of instruction. However, the treatment gains were not significant when measured by a delayed posttest administered four months later.

In one of the few studies that used instruction to study the effects of test-wiseness training on elementary school achievement test scores, Kalechstein et al. (1981) focused on 50 black second grade inner city students. The subjects received ten hours of instruction over five weeks in the following techniques: following directions, marking responses, time using strategies, and guessing strategies. Kalechstein et al. (1981) found significant differences in standardized test scores in favor of the group that received training when controlling for reading ability. Kalechstein et al. (1981) suggested that scores of culturally different students should be interpreted with extreme caution, and that no far reaching conclusions about achievement be drawn without consideration of the problems of lack of test-wiseness skills in these individuals.

More recently, Dreisbach and Keogh (1982) taught four test-taking skills to 131 Spanish speaking kindergarten children from low socioeconomic backgrounds. Half of the children were trained during two 20 minute sessions on four content free skills: familiarization with standardized test vocabulary, response marking, auditory and direction follow-

ing skills, and an introduction to the purpose of testing. The control group spent an equal amount of time in drawing activities. One week after training, each child was tested on the El Circo Readiness Test in both English and Spanish. The results of the study showed that the children trained in the four test-taking skills performed better than the control group in both language tests. Also, based on mean differences the girls performed significantly better than the boys overall.

While significant gains have been made in accumulating knowledge in the area of test-wiseness training, much research remains to be done. To illustrate, there has been a paucity of research on elementary school children, and even less has been done on minority children. Sex differences have been similarly ignored. Another unresolved issue is whether test-wiseness training is equally effective for math and verbal achievement. Finally, the long term effects of test-wiseness training have not been fully explored. Therefore, the purpose of the present study was to extend the test-wiseness literature by investigating the effect of test-wiseness training, ethnicity and sex on the verbal and math achievement of third and fifth grade students.

METHODOLOGY

Sample

A sample of third and fifth grade Black, White and Hispanic students was selected from three public elementary schools in a large urban school district. The sample consisted of 208 students: 37% Hispanic, 43% White, 20% Black; 37% third grade, 63% fifth grade; and 48% males, 52% females.

Instrumentation

Third Grade. The achievement criterion for the third grade students was the reading comprehension and math computation (Part B) score on the Stanford Achievement Test (SAT) (Primary Level III, Form A). The reading comprehension subtest consisted of 70 items, while the math computation test had 12 items. Students were given 35 minutes to complete the reading comprehension test and 10 minutes to finish the math computation section, in addition to the time needed for teacher directions. In addition to the SAT, the students' Comprehensive Test of Basic Skills (CTBS) score was obtained as a delayed posttest measure. The CTBS is administered by the local school district in the late Spring of each year.

Fifth Grade. The score from the vocabulary subtest and the math computation subtest (Part B) on the Stanford Ac-

hievement Test (Intermediate Level II, Form B) served as the criterion measure for the fifth grade subjects. The vocabulary test consisted of 50 items and the math computation test contained 21 items. Students were given approximately 25 minutes to complete the vocabulary test and 20 minutes for the math computation test, including the time necessary to complete the directions. Also, the students' CTBS score was obtained as a delayed posttest measure for the fifth grade.

Materials

A test-wiseness curriculum manual was developed to teach four test-taking strategies: following directions, using time well in taking tests, guessing strategies, and answer changing. These skills have been shown to produce significant improvements in the performance of elementary school children who have had little or no previous test-taking experience on standardized achievement tests (Callenbach, 1973; Dreisbach & Keogh, 1982; Kalechstein, et al., 1981). In developing this manual, steps were taken to avoid teaching skills in which emphasis on knowledge of subject matter was necessary. (A copy of the manual is available from the authors). The reading level of the manual and the skills taught were piloted on a different sample of third and fifth grade students prior to the study.

Procedures

A pretest-posttest-delayed posttest design with a control group was used in the study. The first posttest was the SAT described above and the delayed posttest was the CTBS given by the District one month later as part of their regular testing program. Due to the nature of this research, teacher cooperation was necessary. Therefore, students were selected from schools where teacher and pupil consent were obtained. A total of 12 classrooms were selected from three schools: two classrooms at each grade level. The experimental and control treatments were randomly assigned to intact classes for each grade within each school. Thus, a quasi-experimental manipulation (test-wiseness training versus control) was used. Because classrooms were left intact, the randomization procedure used in this study would not serve as a sufficient safeguard against any initial selection bias; therefore, the pretest was used to identify any pretreatment differences between the groups.

One week prior to the test-wiseness training all children in both the experimental and control groups were pretested using alternate forms of the two subtests of the SAT for each respective grade. All testing was performed during regular school hours by trained volunteer graduate students following standard procedures provided in the test manual. Neither the experimental or control groups were informed as to the true nature of the pretesting. However, each respec-

tive teacher informed the children prior to the pretesting that a group of people would be giving a test to them.

One week after the pretest period the test-wiseness training was performed by volunteer graduate students trained by the investigators in the test-taking skills curriculum. There were a total of five trainers, two of which were present in each classroom during each training period. Each team of trainers was assigned to a third and fifth grade class, and remained with those classes for the duration of the training period. The test-wiseness training was conducted for 30 minutes over three sessions in each third and fifth grade classroom. Two different skills were taught on each of the first two sessions and a review of all four skills was conducted on the third session. The last ten minutes of each training session was devoted to practicing the skills taught for that day. Two days after the final training session the experimental and control groups were given the SAT posttest in verbal comprehension and math computation. The same testing procedures followed during the pretesting session were used during the posttesting period. The delayed posttest was administered one month later by the local school district.

RESULTS AND DISCUSSION

All dependent variables were converted to T scores for ease of comparison across achievement areas, grade levels

and from post to delayed posttesting. The results were analyzed using regression procedures. Separate regressions were computed for each dependent variable and each grade level. Treating each grade separately was necessitated by the fact that the design was not fully crossed (i.e., there were no Black third grade students in the study). The order of entry into the regression was a) sex and race, b) pretest verbal and pretest math (for both of the criterion variables), c) treatment and d) the two interactions of sex and treatment and race and treatment. This ordering of variables was used to control for variance that might artifically influence the two major hypotheses of interest in the study: 1) the interactions of training and ethnicity and training and sex and 2) the main effects of training, ethnicity and sex. Controlling for pretest differences, neither of the two interactions involving training were significant on the posttest or the delayed posttest. Results of the main effects analyses will be reported only where significant differences were found.

Training Effects

Experimental and control group posttest means, adjusted for race, sex and pretests are given in Table 1. The training group scored significantly higher on the math test in the fifth grade ($F(1,118) = 7.83$, $p<.05$). However, no other effects due to training were significant.

The delayed posttest adjusted means are shown in Table 2. The results were analyzed as described above. Inspection of Table 2 indicated that the effect observed on the immediate SAT posttest generalized to the delayed CTBS posttest. That is, the experimental group scored substantially higher than the control group ($F(1,97) = 16.71$, $p<.05$).

Insert Tables 1 and 2 About Here

Sex and Race Main Effects

Only one sex difference across the two grades was found. This difference was in favor of the third grade females ($X = 52.8$), who scored approximately one standard deviation higher on math achievement in the third grade than the males ($X = 47.5$) ($F(1,74) = 5.36$, $p<.05$). Race differences are shown in Table 3 for the posttest. As expected from prior studies, Whites scored significantly higher than Blacks and Hispanics on the two achievement measures at the third grade. Slightly different results were obtained from the delayed posttest regarding race differences (Table 4). For the third grade, the Whites scored significantly higher on the verbal test than the Hispanics, but no differences were noted for math. At the fifth grade, the Whites again scored significantly higher in verbal achievement than the Hispan-

ics and Blacks. However in math achievement, the Hispanics' scores increased so that the White group only scored higher than the Black students whose achievement did not change from the post to the delayed posttest.

Insert Tables 3 & 4 About here

In summary, the test-wiseness training was effective, but its effectiveness was limited to math achievement in the fifth grade sample. Nonetheless, the training effects in math achievement is viewed as an important finding, particularly in light of the fact that it generalized to a delayed posttest given one month later. This finding was quite meaningful since previous researchers have tended to focus their efforts primarily on verbal skills. Thus, test-wiseness training while being a verbally oriented skill, appeared to transfer to a non-veral area and may truly be independent of content as Slakter et al., (1970) had contended. Since the interactions between race and training were not significant, it is not possible to say which ethnic group benefited the most from the training. In addition, the lack of significant findings at the third grade level was attributed in part to the fact that these students had not been apart of the District testing program, and as such did not have much testing experience. Therefore, the train-

ing may not have been as meaningful to them as it was to the older fifth grade students.

CONCLUSIONS AND RECOMMENDATIONS

By the very nature of research, each study while answering some questions, tends to reveal many new areas to investigate. The following recommendations for future research were generated from conflicting and serendipitious findings, classroom observations and limitations of the present study.

1. The test-wiseness training, which in the present study was conducted in three 30 minutes sessions should be expanded in terms of the number of training sessions. It is believed that this would result in more positive transfer of test-wiseness skills to standardized testing situations especially for the lower grade students who may unfamiliar with standardized testing procedures.

2. The test-wiseness training group not only received the training, but also had practice in using those skills on test items that were very similar in format to the pre and posttests. In contrast, the control group did not receive training or practice. Thus, the training was confounded with practice. It is suggested that future research provide a third treatment group that would receive only the practice problems in a standardized testing format as practice may be as effective as the training.

3. It is suggested that research be conducted to determine which of the components of test-wiseness would be most advantageous since most of the research has shown that training on several combined skills can produce an increase in standardized test scores.

4. Future research should continue to be directed at assessing the effects of test-wiseness training in different content areas. If the principles of test-wiseness are independent of content, empirical research needs to be directed at providing validity to what has previously been assumed to be true.

5. The relationship between test-wiseness and various personality variables, such as test anxiety, achievment motivation or attributions regarding testing needs to be investigated. Are motivation and anxiety levels confounded with test-wiseness training? It may be the case that the changes in motivation and anxiety that result from test-wiseness training are the cause of the increase in achievement test scores, and not the training itself.

By understanding why certain student subgroups are unable to use testing strategies to perform well on tests, remediation techniques can be developed to help these students in the future, and thus, improve the educational equity for all students. In addition, understanding how a lack of test-wiseness ability in some students may act to depress achievement will aid educators who must draw implications

and develop educational policy based upon the interpretation of discrepant testing results.

REFERENCES

Anastasi, A., & Cordova, F. Some effects of bilingualism upon the intelligence test performance of Puerto Rican children in New York City. Journal of Educational Psychology, 1953, 44, 1-19.

Ayrer, J., Diamond, J.J., Fishman, R., & Green, P. Are inner city children test-wise? Journal of Educational Measurement, 1976, 14, 39-45.

Bajtelsmit, J. W. Test-wiseness and systematic desensitization programs for increasing test-taking skills. Journal of Educational Measurement, 1977, 14, 335-341.

Benson, J. A redefinition of content validity. Educational and Psychological Measurement, 1980, 41, 793-802.

Benson, J., Urman, H. & Hocevar, D. The relationship between race and test-wiseness in third through sixth grade children. National Institute of Education Final Report, No. G 81-0081, September 1982.

Callenbach, C. The effects of instruction and practice in content-independent test-taking techniques upon the standardized reading test socres of selected second grade students. Journal of Educational Measurement, 1973, 10, 25-30.

Crehan, K.D., Koehler, R.A., & Slakter, M.J. Longitudinal studies of test-wiseness. Journal of Educational Measurement, 1974, 11, 209-212.

Crocker, L., & Benson, J. Does answer changing affect test quality? Measurement and Evaluation in Guidance, 1980, 12, 147-153.

Diamond, J.J., & Evans, W.J. An investigation of the cognitive correlates of test-wiseness. Journal of Educational Measurement, 1972, 9, 145-150.

Dreisbach, M., & Keogh, B.K. Test-wiseness as a factor in readiness test performance of young mexican-american children. Journal of Educational Psychology, 1982, 74, 224-229.

Ebel, R. L. Measuring Educational Achievement. Englewood Cliff, N.J.: Prentice Hall, Inc., 1965.

Ebel, R.L., & Damirin, D.E. Tests and examinations. In C.W. Harris (Ed.), Encyclopedia of Educational Research. (3rd edition) New York: The MacMillan Co., 1960.

Eels, K., David, H., Havighurst, R., Herrick, U., & Tyler, R. Intelligence and cultural differences, University of Chicago Pres , 1951.

Flynn, J. & Anderson, B. The effects of test item cue sensitivity on IQ and achievement test performance. Educational Research Quarterly, 1977, 2, 32-39.

Ford, A. & Weener, P.D. The influence of two test-wiseness programs upons students' test performance. Paper presented at the annual meeting of the American Educational Research Association, Boston, April, 1980.

Gaines, W.G., & Jongsma, E.A. The effect of training in test-taking skills on the achievement scores of fifth grade pupils. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, Illinois, April 1974.

Gibb, B.G. Test-wiseness as a secondary cue response. (Doctoral dissertation, Stanford University) Ann Arbor, Michigan: University Microfilms, 1964. No. 64-7643.

Goslin, E., Epstein, R.R., & Hallock, A. The use of standardized tests in elementary schools. New York: Russell Sage Foundation, 1965.

Gross, L.J. The effects of three selected aspects of test-wiseness on the standardized test performance of eight grade students. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, Calif., April 1976.

Kalechstein, P., Kalechstein, M., & Doctor, R. The effects of instruction on test-taking skills in second grade black children. Measurement and Evaluation in Guidance, 1981, 13, 198-202.

Langer, J., Work, D. & Johnson, S. Test-wiseness in objective tests. In P.L. Nacke (Ed.). Diversity in mature reading: Theory and Research, Vol. 1, 22nd Yearbook of the National Reading Conference. National Reading Conference, Milwaukee, Wis., 1973.

Mc Morris, R.F., Brown, J.A., Snyder, G.W & Pruzek, R.M. Effects of violating item construction principles. Journal of Educational Measurement, 1972, 9, 287-295.

Mercer, M. Answer changing and students' test scores. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1979.

Mercer, J., & Brown, W. Racial differences in I.Q.: Fact or artifact? In Carl Senna (Ed.), The fallacy of I.Q. New York: The Third Press, Joseph Okapaku Publishing Co., Inc., 1973.

Metfessel, N.S., & Sax, G. Systematic biases in the keying of correct responses on certain standardized tests. Educational and Psychological Measurement, 1958, 18, 787-790.

Millman, J., Bishop, C.H., & Ebel, R. An analysis of test-wiseness. Educational and Psychological Measurement, 1965, 25, 707-726.

Moore, J.C. Test-wiseness and analogy test performance. Measurement and Evaluation in Guidance, 1971, 3, 198-202.

Moore, J.C., Shutz, R.E., & Baker, R.L. The application of a self-instructional technique to develop a test-taking strategy. American Educational Research Journal, 1966, 3, 13-17.

Morse, L.W. & Morse, D.T. The effects of teacher style and specificity of instructional materials on learning and retaining different tasks. Paper presented at the annual meeting of the American Educational Research Association, Boston, April, 1980.

Oakland, T. The effects of test-wiseness materials on standardized test performance of preschool disadvantaged children. Journal of School Psychology, 1972, 10, 355-360.

Omvig, G. Effects of guidance on the results of standardized achievement testing. Measurement and Evaluation in Guidance, 1971, 4, 47-52.

Ortar, G. Improving test validity by coaching. Educational Research, 1960, 2, 137-142

Ebel, R. L. Measuring Educational Achievement. Englewood Cliff, N.J.: Prentice Hall, Inc., 1965.

Ebel, R.L., & Damirin, D.E. Tests and examinations. In C.W. Harris (Ed.), Encyclopedia of Educational Research. (3rd edition) New York: The MacMillan Co., 1960.

Eels, K., David, H., Havighurst, R., Herrick, U., & Tyler, R. Intelligence and cultural differences, University of Chicago Press, 1951.

Flynn, J. & Anderson, B. The effects of test item cue sensitivity on IQ and achievement test performance. Educational Research Quarterly, 1977, 2, 32-39.

Ford, A. & Weener, P.D. The influence of two test-wiseness programs upons students' test performance. Paper presented at the annual meeting of the American Educational Research Association, Boston, April, 1980.

Gaines, W.G., & Jongsma, E.A. The effect of training in test-taking skills on the achievement scores of fifth grade pupils. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, Illinois, April 1974.

Gibb, B.G. Test-wiseness as a secondary cue response. (Doctoral dissertation, Stanford University) Ann Arbor, Michigan: University Microfilms, 1964. No. 64-7643.

Goslin, E., Epstein, R.R., & Hallock, A. The use of standardized tests in elementary schools. New York: Russell Sage Foundation, 1965.

Gross, L.J. The effects of three selected aspects of test-wiseness on the standardized test performance of eight grade students. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, Calif., April 1976.

Kalechstein, P., Kalechstein, M., & Doctor, R. The effects of instruction on test-taking skills in second grade black children. Measurement and Evaluation in Guidance, 1981, 13, 198-202.

Langer, G., Work, D. & Johnson, S. Test-wiseness in objective tests. In P.L. Nacke (Ed.). Diversity in mature reading: Theory and Research, Vol. 1, 22nd Yearbook of the National Reading Conference. National Reading Conference, Milwaukee, Wis., 1973.

Mc Morris, R.F., Brown, J.A., Snyder, G.W & Pruzek, R.M. Effects of violating item construction principles. Journal of Educational Measurement, 1972, 9, 287-295.

Mercer, M. Answer changing and students' test scores. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1979.

Mercer, J., & Brown, W. Racial differences in I.Q.: Fact or artifact? In Carl Senna (Ed.), The fallacy of I.Q. New York: The Third Press, Joseph Okapaku Publishing Co., Inc., 1973.

Metfessel, N.S., & Sax, G. Systematic biases in the keying of correct responses on certain standardized tests. Educational and Psychological Measurement, 1958, 18, 787-790.

Millman, J., Bishop, C.H., & Ebel, R. An analysis of test-wiseness. Educational and Psychological Measurement, 1965, 25, 707-726.

Moore, J.C. Test-wiseness and analogy test performance. Measurement and Evaluation in Guidance, 1971, 3, 198-202.

Moore, J.C., Shutz, R.E., & Baker, R.L. The application of a self-instructional technique to develop a test-taking strategy. American Educational Research Journal, 1966, 3, 13-17.

Morse, L.W. & Morse, D.T. The effects of teacher style and specificity of instructional materials on learning and retaining different tasks. Paper presented at the annual meeting of the American Educational Research Association, Boston, April, 1980.

Oakland, T. The effects of test-wiseness materials on standardized test performance of preschool disadvantaged children. Journal of School Psychology, 1972, 10, 355-360.

Omvig, G. Effects of guidance on the results of standardized achievement testing. Measurement and Evaluation in Guidance, 1971, 4, 47-52.

Ortar, G. Improving test validity by coaching. Educational Research, 1960, 2, 137-142

Pike, L.W. Implicit guessing of GRE-aptitude examinees classified by ethnic group and sex (GRE Board Professional Report GREB No. 75). Princeton, N.J.: Educational Testing Service, 1980.

Pike, L.W. Short-term instruction, test-wiseness and Scholastic Aptitude Test: A literature review with research recommendations (CB RDR 77-78, No. 2 and ETS RB78-2). Princeton, N.J.: Educational Testing Service, 1978.

Powers, R.S., & Sabers, S.A. An investigation of ethnic group differences in test-wiseness at the third, fifth and seventh grade. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April, 1981.

Sarnacki, F.E. An examination of test-wiseness in the cognitive test domain. Review of Educational Research, Spring, 1979, 4, 252-279.

Schmitt and Crocker, 1981 Improving examinee performance on multiple-choice tests. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April, 1981.

Slakter, M. J. The penalty for not guessing. Journal of Educational Measurement. 1968, 5, 141-144. (a)

Slakter, M. J. The effect of guessing strategy on objective test scores. Journal of Educational Measurement, 1968, 5, 217-222. (b)

Slakter,M.J., Koehler, R.A., & Hampton, S.H. Grade level, sex and selected aspects of test-wiseness. Journal of Educational Measurement, 1970, 7, 119-122. (a)

Slakter, M.J., Koehler, R.A., & Hampton, S.H. Learning test-wiseness by programmed texts. Journal of Educational Measurement, 1970, 7, 247-254. (b)

Stanley, J. C. Reliability. In R.L. Thorndike (Ed.). Educational Measurement. Washington, D.C.: American Council on Education, 1971.

Thorndike, E.L. Reliability. In E.F. Lindquist (Ed.), Educational Measurement. Washington, D.C.: American Council on Education, 1951.

Wahlstrom, M., & Boersma, F.J. The influence of test-wiseness upon achievement. Educational and Psychological Measurment, 1968, 28, 413-420.

Woo, T., Hocevar, D. & Benson, J. Grade level and home language effects upon impulsivity manifestation on standardized reading achievement. Manuscript in preparation, 1983.

Woodley, K.K. Test-wiseness program development and evaluation. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, 1973.

Woodley, K.K. Test-wiseness: A cognitive function? Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, D.C., April, 1975.

TABLE 1

Adjusted Posttest Means and F ratios

Training vs. Control Groups

N=208

	Treatment		
Grade/Test	Test-wiseness	Control	F ratio
Third (N=76)	(n=44)	(n=32)	
Verbal	50.6	49.0	1.60
Math	50.1	51.0	<1.00
Fifth (N=132)	(n=64)	(n=68)	
Verbal	50.1	49.9	<1.00
Math	51.7	48.4	7.83*

Note. The means were adjusted by sex, ethnicity and the verbal and math pretests. The effect size for math was .33
*$p < .05$

TABLE 2

Adjusted Delayed Posttest Means and F ratios

Training vs. Control Groups

N=176

	Treatment		
Grade/Test	Test-wiseness	Control	F ratio
Third (N=66)	(n=40)	(n=26)	
Verbal	49.4	52.3	1.74
Math	50.6	50.2	<1.00
Fifth (N=111)	(n=60)	(n=51)	
Verbal	50.7	49.2	1.76
Math	52.7	46.8	16.71*

Note. The means were adjusted by sex, ethnicity and the verbal and math pretests. The effect size for math was .59.
*$p<.05$

Table 3

Adjusted Posttest Means and F ratios

Race Differences In Achievement

N=208

	Ethnicity			
Grade/Test	White	Black	Hispanic	F ratio
Third (N=76)	(n=30)	(n=0)	(n=46)	
Verbal	56.1	-----	45.9	24.7*
Math	53.7	-----	48.4	5.4*
Fifth (N=132)	(n=45)	(n=37)	(n=50)	
Verbal	57.3	49.1	45.0	31.8*
Math	55.8	46.4	47.9	13.6*

Note. The means are adjusted by sex differences.

*$p < .05$

Table 4

Adjusted Delayed Posttest Means and F ratios

Race Differences In Achievement

N=177

	Ethnicity			
Grade/Test	White	Black	Hispanic	F ratio
Third (N=66)	(n=24)	(n=0)	(n=42)	
Verbal	56.5	-----	46.9	17.4*
Math	53.2	-----	48.8	3.1
Fifth (N=111)	(n=43)	(n=34)	(n=34)	
Verbal	56.0	46.3	46.2	16.8*
Math	53.0	46.9	49.3	4.3*

Note. The means are adjusted by sex differences.

*$p < .05$